

CLAIMS

Having thus described my invention, what I claim as new and desire to secure by Letters Patent is as follows:

1. A method of conducting a survey, said method comprising:
for at least one question in said survey, establishing a bin for each of a possible response to said question; and
for each said bin, establishing a perturbing mechanism that perturbs a content of said bin, said perturbing mechanism having a statistical parameter with a known value.
2. The method of claim 1, further comprising:
generating a perturbed indicator vector that represents a respondent's response for said question, said perturbed indicator vector comprising an information structure including the contents of all bins of said question after each of the bins has been perturbed and said respondent has selected one or more said possible responses.
3. The method of claim 1, wherein said perturbing mechanism comprises a random number generator and said known statistical parameter value comprises a mean value of said random number generator.

4. The method of claim 2, wherein said generating the perturbed indicator vector comprises respectively adding numbers from the perturbing mechanism to the contents of the bins.

5. The method of claim 4, further comprising:

if contents of a bin exceeds an upper bound after perturbation, said contents are clamped to said upper bound.

6. The method of claim 4, further comprising:

if contents of a bins are below a lower bound after perturbation, said contents are clamped to said lower bound.

7. The method of claim 1, further comprising at least one of:

setting up a survey question by generating a medium with a plurality of markable areas for each possible response and pre-marking a random number of said markable areas for each said possible response; and

having a respondent respond to the survey question by adding a mark to any of remaining non pre-marked markable areas, if any markable areas remain after said pre-marking, of the plurality of markable areas for the possible response that corresponds to a desired response to the question.

8. The method of claim 7, further comprising:

generating a perturbed indicator vector by counting the number of marked areas for each response.

9. The method of claim 2, further comprising:

for a plurality of responses for a question in said survey, analyzing the bins in said perturbed indicator vector to provide an estimation of a distribution of responses.

10. The method of claim 9, wherein said analyzing comprises:

for said question being analyzed, calculating an average of each perturbed bin in said question.

11. The method of claim 10, wherein said perturbing mechanism comprises a random number generator and said known statistical parameter comprises a mean value, said analyzing further comprising:

for each said perturbed bin in said question, subtracting said mean value of said perturbing mechanism associated with said bin.

12. The method of claim 1, further comprising:

for each perturbed bin in said question, adjusting a content of said perturbed bin by an amount of said known value of said statistical parameter.

13. An apparatus for conducting a survey, said apparatus comprising at least one of:

a memory for storing a plurality of respondents' responses to a question in said survey, wherein each of said respondents' response comprises a bin for each of a possible response to said question and a value of each said bin has been perturbed by a perturbing mechanism;

a survey question set-up module to allow a question in said survey to be set up, said survey question set-up module including a module to establish a bin for each of a possible response to said question and a perturbing mechanism for each said bin that perturbs a content of said bin, each said perturbing mechanism having a known value for a statistical parameter;

a respondent module including a selection module to allow a respondent to select at least one of said possible responses to said question and a module to apply said perturbing mechanism for each said bin to generate a perturbed indicator vector corresponding to said respondent's selection;

an analysis module to retrieve data from said database for said question and to analyze each said bin; and

a graphic user interface to allow a user to interface with at least one of said memory, said survey question set-up module, said respondent module, and said analysis module.

14. The apparatus of claim 13, wherein said memory stores a running sum of the perturbed indicator vectors.

15. The apparatus of claim 13, wherein said analysis module calculates an average value of each said bin of said question by dividing a sum of the contents in said bin of the perturbed indicator vectors by the number of respondents.

16. A system for conducting a survey, comprising at least one of:

a memory means for serving as a database to store a plurality of respondent's responses to a question in said survey, wherein each said response comprises a plurality of bins corresponding to a number of possible answers for said question and each said bin is perturbed in value by a perturbing mechanism;

a survey set-up means for setting up a question in said survey, wherein said setting up said question comprises establishing a bin for each of a possible response to said question and establishing a perturbing mechanism that perturbs a content of said bin, said perturbing mechanism having a statistical parameter with a known value;

a respondent means for allowing a respondent to select at least one of said possible answers to said question, for perturbing a content of each bin in said question upon completion of the selection by said respondent, for generating a perturbed indicator vector that includes the contents of all said bins in said question after perturbation, and for transmitting said perturbed indicator vector to said database; and

an analysis means for retrieving and analyzing a content of said bins; and

a user interface means for allowing a user to interface with at least one of said memory means, said survey set-up means, said respondent means, and said analysis means.

17. A system for conducting a survey as in claim 16, wherein said memory means stores a running sum of the perturbed indicator vectors.

18. A system for conducting a survey as in claim 16, wherein said analysis means calculates an average value for said bin by dividing, by the number of respondents being analyzed, the sum of the contents of said bin for all said respondents being analyzed.

19. A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method of at least one of conducting, processing, and analyzing a survey, said program comprising at least one of:

a memory interface to interface with a database that stores a plurality of responses to a question in said survey, wherein each said response comprises a perturbed vector comprising a bin for each of a possible response for said question, as perturbed by a perturbing mechanism having a known value for a statistical parameter;

a survey set-up module to allow a question in said survey to be set up, wherein said set-up module establishes a bin for each of a possible response to said question, and establishes a perturbing mechanism to perturb a content of said bin, said perturbing mechanism having a known value for a statistical parameter;

a respondent module to allow a respondent to select at least one of said possible responses to said question, to perturb a content of each bin in said question upon completion of the selection, to generate a perturbed indicator vector that includes the

contents of all said bins in said question after perturbation, and to send said perturbed indicator vector to said database;

an analysis module to retrieve from said database and analyze data for said question; and

a graphic user interface to allow a user to interface with at least one of said memory device, said survey set-up module, said respondent module, and said analysis module.

20. The signal-bearing medium of claim 19, wherein said memory interface includes instructions to store a running sum of the perturbed indicator vectors.

21. The signal-bearing medium of claim 20, wherein said analysis module calculates an average value by dividing the sum of the contents of a bin of said perturbed indicator vectors by the number of respondents being analyzed.

22. A business method, comprising at least one of:

preparing a survey question in a manner such that, for at least one question in said survey, establishes a bin for each of a possible response to said question, for each said bin, establishes a perturbing mechanism that perturbs a content of said bin, said perturbing mechanism having a known value for a statistical parameter;

allowing users to respond to said survey question;

at least one of receiving and storing said survey question;

transmitting a perturbed indicator vector of a respondent's response to a survey question prepared in said manner, said perturbed indicator vector comprising an information structure including the contents of all bins of said question after each of the bins has been perturbed;

at least one of receiving said perturbed indicator vector and storing said perturbed indicator vector in a database;

at least one of retrieving and analyzing data for said survey question to provide a result of said survey; and

at least one of transmitting, receiving, printing out, and receiving a printed copy of said result.

23. A method of conducting a survey, said method comprising:

for at least one question in said survey, generating an indicator vector from a vector whose components respectively represent a possible response to said question, said indicator vector indicating which of said possible responses were selected by a respondent; and

adding a perturbation vector to said indicator vector to provide a perturbed indicator vector,

said perturbation vector having a same number of components as said indicator vector, each component in said perturbation vector resulting from a perturbation mechanism that is independent of the perturbation mechanism of the other components, wherein said perturbation mechanism has a statistical parameter whose value is known.

24. The method of claim 23, wherein said perturbation mechanism comprises a random number generator and said statistical parameter comprises a mean, said method further comprising:

for each at least one question in said survey, calculating an average perturbed indicator vector from a plurality N of perturbed indicator vectors, wherein vector components of said average perturbed indicator vector respectively comprise an average value of corresponding components in said plurality N of perturbed indicator vectors; and

subtracting from said average perturbed indicator vector a mean vector whose components respectively comprise said mean value of said perturbation mechanism for each said indicator vector component.

25. A method of privacy-preserving data mining, comprising:

for at least one question in a survey used in said data mining, generating an indicator vector from a vector whose components represent each possible response to said question, said indicator vector indicating which of said possible responses were selected by a respondent; and

adding a perturbation vector to said indicator vector to provide a perturbed indicator vector, said perturbation vector having a same number of components as said indicator vector, each component in said perturbation vector resulting from a perturbation mechanism that is independent of the perturbation mechanism of the other components, wherein each said perturbation mechanism has a statistical parameter with a value that is known.

26. A data mining apparatus, comprising:

an indicator vector generator to generate an indicator vector representing a response by a respondent to a survey question;

a perturbation vector generator to generate a perturbation vector; and

a perturbed indicator vector generator to add said indicator vector with said perturbation vector,

wherein, for said question, a predefined possible-response vector exists whose components respectively represent a possible response to said question, said indicator vector comprising a modification of said possible-response vector that represents which one or ones of said possible responses were selected by a respondent, said perturbation vector comprising a vector having a same number of components as said indicator vector, each component in said perturbation vector resulting from a perturbation mechanism that is independent of the perturbation mechanism of the other components, each said perturbation mechanism having a statistical parameter with a value that is known.

27. A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method of conducting a survey, said program comprising:

an indicator vector generator to generate an indicator vector representing a response by a respondent to a survey question;

a perturbation vector generator to generate a perturbation vector; and

a perturbed indicator vector generator to combine said indicator vector with said perturbation vector,

wherein, for each said question, a possible-response vector exists whose components respectively represent a possible response to said question, said indicator vector comprising a modification of said possible-response vector that represents which of said possible responses were selected by a respondent, said perturbation vector comprising a vector having a same number of components as said indicator vector, each component in said perturbation vector resulting from a perturbation mechanism that is independent of the perturbation mechanism of the other components, each perturbation mechanism having a statistical parameter with a known value.